Raw Filtering for IoT Applications on FPGAs

Big Data Applications often work with unstructured or semi-structured data-formats such as JSON. Applications can spend 80%-90% of their execution time on parsing these files to determine records and their attributes. After parsing, execution-pipelines often contain filters, which check the parsed records for filter conditions and can thus reduce the number of records passed on to the following stages. The idea of raw filtering is to apply approximate filters already on the incoming raw bytestream to filter out irrelevant records before they are parsed.

In this work, different concepts for approximate filtering of various IoT workloads will be investigated. Besides string and number comparisons, also id lookups or geofences will be considered as filter candidates. Since raw filtering is based on ASCII characters, special concepts must be developed for this purpose. These will then be implemented for an FPGA and evaluated for resource consumption and effectiveness on benchmarks.

Prerequisites: Basic knowledge in C++, Python and VHDL
Type of Work: Theory (20%), Conception (50%), Implementation (30%)
Supervisor: Tobias Hahn (tobias.hahn@fau.de)